

CONTINGENCY TABLES WITH UNIFORMLY BOUNDED ENTRIES

AUSTIN SHAPIRO

ABSTRACT. We consider nonnegative integer matrices with specified row and column sums and upper bounds on the entries. We show that the logarithm of the number of such matrices is approximated by a concave function of the row and column sums. We give efficiently computable estimators for this function, including one suggested by a maximum-entropy random model; we show that these estimators are asymptotically exact as the dimension of the matrices goes to ∞ . We finish by showing that, for $\kappa \geq 2$ and for sufficiently small row and column sums, the number of matrices with these row and column sums and with entries $\leq \kappa$ is greater by an exponential factor than predicted by a heuristic of independence.

1. MAIN OBJECTS AND RESULTS

A *contingency table* is defined as a nonnegative integer matrix with specified row and column sums. Specifically, given

$$R = (r_1, r_2, \dots, r_m) \in \mathbb{Z}_{\geq 0}^m \quad \text{and} \quad C = (c_1, c_2, \dots, c_n) \in \mathbb{Z}_{\geq 0}^n$$

such that

$$r_1 + r_2 + \dots + r_m = c_1 + c_2 + \dots + c_n = N,$$

we denote by $\Pi(R, C)$ the set of all $A = (a_{ij}) \in \mathbb{R}_{\geq 0}^{m \times n}$ such that

$$\sum_{j=1}^n a_{ij} = r_i \quad (1 \leq i \leq m) \quad \text{and} \quad \sum_{i=1}^m a_{ij} = c_j \quad (1 \leq j \leq n).$$

Then $\Pi(R, C)$ is a convex polytope, and its integer points are known as the contingency tables with *margins* R and C ; we denote the number of such tables by $T(R, C)$.

Given a matrix $K = (k_{ij}) \in (\mathbb{Z}_{\geq 0} \cup \{\infty\})^{m \times n}$, define

$$\Pi_K(R, C) := \{A \in \Pi(R, C) : a_{ij} \leq k_{ij} \quad \forall i, j\}.$$

Thus the integer points of $\Pi_K(R, C)$ are the contingency tables with margins R and C , bounded entrywise by K . We denote the number of such tables by $T_K(R, C)$.

Date: February 2011.

The author was supported in part by NSF grants DMS 0400617 and DMS 0856640.

We will be particularly concerned with the case in which K is constant, with all entries equal to $\kappa \in \mathbb{Z}_{>0} \cup \{\infty\}$, in which case we abuse the notation slightly by writing $\Pi_\kappa(R, C)$ in lieu of $\Pi_K(R, C)$; $T_\kappa(R, C)$ in lieu of $T_K(R, C)$; and so on. The integer points of $\Pi_\kappa(R, C)$ are contingency tables with margins R and C and entries restricted to $\{0, 1, 2, \dots, \kappa\}$. When $\kappa = 1$, these tables are known as *0–1 tables*. When $\kappa = \infty$, it is to be understood that $\{0, 1, 2, \dots, \kappa\}$ signifies $\mathbb{Z}_{\geq 0}$, so that $\Pi_\infty(R, C) := \Pi(R, C)$.

Contingency tables arise in statistics, where they are used to represent the joint distribution of two categorical variables in a sample. In this context, one may wish to test how significantly an observed table A deviates from a “typical” table with the observed margins. A plausible candidate for the “typical” table is the *rank 1 table*,

$$A^{\text{ind}} := \left(\frac{r_i c_j}{N} \right)_{i,j}.$$

Deviation from this table tests a hypothesis of independence between the two categorical variables.

In a similar spirit, Good [12] proposed the following heuristic for estimating $T(R, C)$: Consider the set of $m \times n$ nonnegative integer matrices with sum of entries equal to N ; there are $\binom{N+mn-1}{mn-1}$ such matrices. Equip this set with the uniform probability measure. Then the probability that a random sample from this set has row margin R is

$$\binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^m \binom{r_i+n-1}{n-1},$$

while the probability that a random sample has column margin C is

$$\binom{N+mn-1}{mn-1}^{-1} \prod_{j=1}^n \binom{c_j+m-1}{m-1}.$$

If these two events were independent, then the number of tables satisfying both constraints would be

$$I(R, C) := \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^m \binom{r_i+n-1}{n-1} \prod_{j=1}^n \binom{c_j+m-1}{m-1}.$$

In [8], see also [9], Diaconis and Efron criticized the choice of A^{ind} as typical, and discussed alternatives to the independence hypothesis, including a hypothesis that A is drawn uniformly at random from the set of tables with the observed margins. This creates the need for a different notion of the “typical” table, which, it turns out, may diverge dramatically from the rank 1 table (see [5]).

Using the theory of permanents and matrix scaling, Barvinok estimated $T(R, C)$ [2] and $T_1(R, C)$ [4] and showed that $T(R, C)$ is approximately log-concave [1]. Barvinok also showed [3] that, in a certain asymptotic sense (see Definition 2, “Cloning”),

$T(R, C)$ is almost always severely underestimated by $I(R, C)$. That is, for most values of R and C , when m and n are large, the conditions

$$\sum_{j=1}^n a_{ij} = r_i \quad (1 \leq i \leq m) \quad \text{and} \quad \sum_{i=1}^m a_{ij} = c_j \quad (1 \leq j \leq n)$$

are strongly positively correlated. Yet if we define

$$I_1(R, C) := \binom{mn}{N}^{-1} \prod_{i=1}^m \binom{n}{r_i} \prod_{j=1}^n \binom{m}{c_j}$$

by analogy to $I(R, C)$, then it turns out that $I_1(R, C)$ typically *overestimates* $T_1(R, C)$; that is, for 0–1 tables, the conditions

$$\sum_{j=1}^n a_{ij} = r_i \quad (1 \leq i \leq m) \quad \text{and} \quad \sum_{i=1}^m a_{ij} = c_j \quad (1 \leq j \leq n)$$

are strongly *negatively* correlated [4] for most values of R and C .

The aim of this paper is to extend these results to the case of general κ , and in particular to examine the transition between the positive and negative correlations described above. Although asymptotic positive correlation does not occur for any margins (R, C) when $\kappa = 1$, we will show that it does occur for *some* margins whenever $\kappa \geq 2$.

1.1. How to read this paper. The four principal theorems of the paper are stated in sections 1.2 and 1.5. Each is proven in its own section (2.1, 2.4, 3.2, 3.4). The intervening sections are thematic, each introducing a concept which will be used in the statements and proofs of the four theorems. These sections contain definitions, lemmas, and proofs of lemmas, as well as informal motivating remarks; they can be read linearly as a “story,” or merely scanned for their formal content. The reader in a hurry is advised to browse these sections for their definitions, then proceed directly to the proof sections, referring back to the thematic sections as needed (e.g., for the statement of lemmas cited in the proofs).

1.2. Results: part 1. We begin by adapting a “Brunn–Minkowski-type” result of Barvinok [1] to prove approximate log-concavity for $T_K(R, C)$. Using the following definitions:

Definition 1. For a vector or matrix V , let $|V|$ denote the sum of the entries of V . For an integer $n \geq 0$, let $\omega(n) := \frac{n^n}{n!}$ (agreeing that $0^0 = 1$). For a vector or matrix V with nonnegative integer entries, let $\Omega(V)$ denote the sum of $\omega(v)$ over all entries v of V .

we can state:

Theorem 1. *Let $\alpha_1 + \alpha_2 + \cdots + \alpha_p = 1$ ($\alpha_1, \dots, \alpha_p \geq 0$). Let $R^1, \dots, R^p \in \mathbb{Z}_{\geq 0}^m$ and $C^1, \dots, C^p \in \mathbb{Z}_{\geq 0}^n$, such that $|R^1| = \cdots = |R^p| = |C^1| = \cdots = |C^p| = N$. Let $K \in \mathbb{Z}_{\geq 0}^{m \times n}$. Define*

$$R := \alpha_1 R^1 + \alpha_2 R^2 + \cdots + \alpha_p R^p \quad \text{and} \quad C := \alpha_1 C^1 + \alpha_2 C^2 + \cdots + \alpha_p C^p.$$

Also, define vectors $\tilde{C}^1, \tilde{C}^2, \dots, \tilde{C}^p, \tilde{C} \in \mathbb{Z}_{\geq 0}^n$ by

$$\tilde{c}_j^t := \left(\sum_{i=1}^m k_{ij} \right) - c_j^t \quad (1 \leq t \leq p, \ 1 \leq j \leq n)$$

and

$$\tilde{c}_j := \left(\sum_{i=1}^m k_{ij} \right) - c_j \quad (= \alpha_1 \tilde{c}_j^1 + \cdots + \alpha_p \tilde{c}_j^p).$$

(The coordinates of vectors with uppercase names are indicated by lowercase letters: e.g., $\tilde{C}^t = (\tilde{c}_1^t, \dots, \tilde{c}_n^t)$.)

Then

$$\frac{\omega(|K|)T_K(R, C)}{\Omega(R)\Omega(\tilde{C})\Omega(K)} \geq \prod_{t=1}^p \left[\frac{T_K(R^t, C^t)}{\min\{\Omega(R^t)\Omega(\tilde{C}^t), \Omega(K)\}} \right]^{\alpha_t}.$$

This theorem is somewhat opaque in itself, due to the confounding factors $\Omega(R^t)$, $\Omega(\tilde{C}^t)$, etc. However, some analysis reveals that these factors typically grow more slowly than the numbers T_K . For a precise statement, we follow [3] and introduce the following asymptotic regime.

Definition 2 (Cloning). *Let*

$$R = (r_1, \dots, r_m) \in \mathbb{Z}_{\geq 0}^m \quad \text{and} \quad C = (c_1, \dots, c_n) \in \mathbb{Z}_{\geq 0}^n.$$

Then we define

$$R^{(s)} = (sr_1, \dots, sr_m, \ sr_1, \dots, sr_m, \ \dots, \ sr_1, \dots, sr_m)$$

and

$$C^{(s)} = (sc_1, \dots, sc_n, \ sc_1, \dots, sc_n, \ \dots, \ sc_1, \dots, sc_n),$$

where the number of repetitions is s (thus $R^{(s)} \in \mathbb{Z}_{\geq 0}^{sm}$ and $C^{(s)} \in \mathbb{Z}_{\geq 0}^{sn}$). We refer to these vectors as the s -fold clonings of R and C .

If $K \in \mathbb{Z}_{\geq 0}^{m \times n}$, then we define $K^{(s)}$ as the $sm \times sn$ matrix of form

$$\begin{pmatrix} K & K & \cdots & K \\ K & K & \cdots & K \\ \vdots & \vdots & \ddots & \vdots \\ K & K & \cdots & K \end{pmatrix}$$

(with s blocks in either direction). We call this the s -fold cloning of K .

Note that the clonings are defined so that, if A is a contingency table with margins R and C , then the $sm \times sn$ matrix

$$\begin{pmatrix} A & A & \cdots & A \\ A & A & \cdots & A \\ \vdots & \vdots & \ddots & \vdots \\ A & A & \cdots & A \end{pmatrix}$$

has margins $R^{(s)}$ and $C^{(s)}$.

Now we can state

Theorem 2. *Let $R \in \mathbb{Z}_{>0}^m$, $C \in \mathbb{Z}_{>0}^n$, and $K \in \mathbb{Z}_{>0}^{m \times n}$. Assume that $T_K(R, C) > 0$, that is, there is at least one contingency table with margins R and C , bounded entrywise by K . Then*

$$\lim_{s \rightarrow \infty} \frac{1}{s^2} \ln T_{K^{(s)}}(R^{(s)}, C^{(s)}) = \ln \left(\inf_{x_1, \dots, x_m, y_1, \dots, y_n > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right),$$

where

$$G(\mathbf{x}, \mathbf{y}) := \prod_{i=1}^m \prod_{j=1}^n [1 + x_i y_j + (x_i y_j)^2 + \cdots + (x_i y_j)^{k_{ij}}]$$

(and $\mathbf{x}^R, \mathbf{y}^C$ denote $x_1^{r_1} x_2^{r_2} \cdots x_m^{r_m}$, $y_1^{c_1} y_2^{c_2} \cdots y_n^{c_n}$).

To turn this into a more usable estimate, we employ some concepts from probability theory.

1.3. Maximum-entropy models. How might we “approximate” the uniform distribution on $\Pi = \Pi_K(R, C)$? One naïve approach is to construct a random matrix with entries bounded by K which satisfies the row and column constraints *on average*, that is, in expectation. Among all possible distributions for this random model, one compelling choice is that which achieves the maximum entropy. Such a distribution necessarily assigns equal mass to all *bona fide* integer points of Π , while also awarding some mass to impostors outside Π . Thus the entropy of the random model overestimates the entropy of the uniform distribution on $\Pi \cap \mathbb{Z}^{m \times n}$, and so provides an upper bound on $\ln T_K(R, C)$. Theorem 3 expresses that this upper bound is actually a tolerably good estimate. The great advantage of the random model is that it is readily computable, with all coordinates being mutually independent.

The reader interested in the general applicability of this method to integer points of polytopes is directed to [6], [16]. The underlying “maximum-entropy principle” has a long history (see for example [14], [15], [13]), and we do not address its overall merits here; it is not a formal prerequisite for any of the results to follow or for their proofs, but it does lend them intuitive plausibility.

The following definitions *are* prerequisites for a statement of the results.

Definition 3. A random variable X is truncated geometric with support $\{0, 1, 2, \dots, \kappa\}$ if there are parameters $p \in (0, 1]$ and $q \in [0, \infty)$, such that

$$\Pr[X = t] = pq^t \quad \text{for } t = 0, 1, \dots, \kappa.$$

For symmetry, we also say that X is truncated geometric with parameters $p = 0$ and $q = \infty$ if $\Pr[X = \kappa] = 1$; however, in what follows, explicit treatment of this case will sometimes be left to the reader.

When $\kappa = \infty$, we have already indicated that $\{0, 1, 2, \dots, \kappa\}$ is to be interpreted as $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$. A random variable on this support is geometric if there are parameters $p \in (0, 1]$ and $q \in [0, 1)$ (in this case necessarily satisfying $p + q = 1$), such that

$$\Pr[X = t] = pq^t \quad \text{for } t = 0, 1, 2, \dots$$

To avoid unnecessary duplication of results, we regard this as a special case of the truncated geometric distribution.

Given $\kappa \in \mathbb{Z}_{\geq 0}$ and $x \in [0, \kappa]$, or $\kappa = \infty$ and $x \in [0, \infty)$, there is a unique truncated geometric distribution with support $\{0, 1, 2, \dots, \kappa\}$ and expected value equal to x .

Definition 4. We denote this distribution by $TG(x; \kappa)$, and its parameters (as in Definition 3) by $p(x; \kappa)$ and $q(x; \kappa)$.

The parameters $p = p(x; \kappa)$ and $q = q(x; \kappa)$ are given implicitly by the equations

$$1 = p(1 + q + q^2 + \dots + q^\kappa), \tag{1}$$

$$x = p(q + 2q^2 + \dots + \kappa q^\kappa), \tag{2}$$

which, to the author's knowledge, cannot be neatly solved in general (but see section 3 for a discussion of the simplest cases, $\kappa = 1$ and $\kappa = \infty$).

Recall that the entropy of a discrete random variable X (here normalized to base e) is given by

$$H[X] := - \sum_{x \in \text{supp}(X)} \Pr[X = x] \ln \Pr[X = x].$$

Among all probability distributions with support in $\{0, 1, 2, \dots, \kappa\}$ and given expectation x , the greatest entropy is achieved by $TG(x; \kappa)$, as is well-known (and may be readily proved by the method of Lagrange multipliers). Hence

Definition 5. Given $\kappa \in \mathbb{Z}_{\geq 0}$ and $x \in [0, \kappa]$, or $\kappa = \infty$ and $x \in [0, \infty)$, let $H_\kappa^{\max}(x)$ denote the entropy of $TG(x; \kappa)$.

We regard H_κ^{\max} as a function on $[0, \kappa]$. We break off a discussion of its formula and basic properties into section 3, having said all that is needed about H_κ^{\max} in order to state the paper's remaining main results.

1.4. The independence heuristic for $\Pi_\kappa(R, C)$. Inspired by Good's estimate $I(R, C)$, we propose and consider an estimate $I_\kappa(R, C)$ for the number of contingency tables in $\Pi_\kappa(R, C)$ when $1 \leq \kappa < \infty$. (Technically, $\kappa = 0$ is allowable in all that follows, though the consequences are trivial.)

Definition 6 (“ $(\kappa + 1)$ -nomial coefficients” [11]). *Let κ be a positive integer. For integers $n \geq 0$ and $0 \leq r \leq n\kappa$, we denote by $\binom{n}{r}_\kappa$ the coefficient of x^r in the polynomial expansion of $(1 + x + x^2 + \cdots + x^\kappa)^n$.*

For integers $n \geq 0$, $r \geq 0$, we define $\binom{n}{r}_\infty$ to be the coefficient of x^r in the power series expansion of $(1 + x + x^2 + \cdots)^n$.

Note that $\binom{n}{r}_1 = \binom{n}{r}$ and $\binom{n}{r}_\infty = \binom{r+n-1}{r} = \binom{r+n-1}{n-1}$. For $\kappa \neq 1, \infty$, there is (to the author's knowledge) no comparably neat exact formula for $\binom{n}{r}_\kappa$.

Consider the uniform probability measure on the set of $m \times n$ matrices with entries in $\{0, 1, 2, \dots, \kappa\}$ and sum of entries equal to N . The number of such matrices is $\binom{mn}{N}_\kappa$. The probability that a random sample from this set has row margin R is

$$\binom{mn}{N}_\kappa^{-1} \prod_{i=1}^m \binom{n}{r_i}_\kappa;$$

the probability that a random sample has column margin C is

$$\binom{mn}{N}_\kappa^{-1} \prod_{j=1}^n \binom{m}{c_j}_\kappa.$$

Thus if these two events were independent, then the number of tables satisfying both constraints would be given by

Definition 7.

$$I_\kappa(R, C) := \binom{mn}{N}_\kappa^{-1} \prod_{i=1}^m \binom{n}{r_i}_\kappa \prod_{j=1}^n \binom{m}{c_j}_\kappa.$$

Just as $T_\kappa(R, C)$ specializes to $T(R, C)$ when $\kappa = \infty$, note that $I_\infty(R, C) = I(R, C)$.

1.5. Results: part 2. We show the following “log-asymptotic” formulas for $T_\kappa(R, C)$ and $I_\kappa(R, C)$:

Theorem 3. *Let $R \in \mathbb{Z}_{\geq 0}^m$, $C \in \mathbb{Z}_{\geq 0}^n$, and $\kappa \in \mathbb{Z}_{>0} \cup \{\infty\}$. Assume $|R| = |C| = N$, and assume that $T_\kappa(R, C) > 0$. Recall the definition of the cloned margins $R^{(s)}, C^{(s)}$ (Definition 2). Then:*

(i)

$$\lim_{s \rightarrow \infty} \frac{1}{s^2} \ln T_\kappa(R^{(s)}, C^{(s)}) = \max_{Z \in \Pi_\kappa(R, C)} \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max}(z_{ij}).$$

(ii)

$$\lim_{s \rightarrow \infty} \frac{1}{s^2} \ln I_\kappa(R^{(s)}, C^{(s)}) = -mnH_\kappa^{\max}\left(\frac{N}{mn}\right) + n \sum_{i=1}^m H_\kappa^{\max}\left(\frac{r_i}{n}\right) + m \sum_{j=1}^n H_\kappa^{\max}\left(\frac{c_j}{m}\right).$$

We pause to note that the right-hand side of Theorem 3(i) is efficiently computable, as it is the maximum of a strictly concave function over a convex polytope; see Lemma 5. This expression represents the entropy of the “random model” described at the beginning of section 1.3.

Theorem 3 equips us to prove the following positive correlation result:

Theorem 4. *Continue the assumptions of Theorem 3; further, suppose $\kappa \geq 2$. Then there exists $\delta = \delta(\kappa) \in (0, 1)$, such that if (R, C) satisfy*

$$\left(\max_{1 \leq i \leq m} r_i\right) \left(\max_{1 \leq j \leq n} c_j\right) < \delta \kappa N$$

then

$$\lim_{s \rightarrow \infty} \frac{1}{s^2} \ln T_\kappa(R^{(s)}, C^{(s)}) \geq \lim_{s \rightarrow \infty} \frac{1}{s^2} \ln I_\kappa(R^{(s)}, C^{(s)}),$$

with strict inequality if neither R nor C is a constant vector (i.e., if it is not the case that $r_1 = \dots = r_m$ or $c_1 = \dots = c_n$).

We do not present a corresponding negative correlation result. Some commentary on the prospects for such a result can be found in section 3.5.

2. APPROXIMATE LOG-CONCAVITY FOR $T_K(R, C)$ AND ITS CONSEQUENCES

The following “Brunn–Minkowski-type inequality” is proven in [1]:

Theorem 5 (Barvinok). *Let $\alpha_1 + \alpha_2 + \dots + \alpha_p = 1$ ($\alpha_1, \dots, \alpha_p \geq 0$). Let $R^1, \dots, R^p \in \mathbb{Z}_{\geq 0}^m$ and $C^1, \dots, C^p \in \mathbb{Z}_{\geq 0}^n$, such that $|R^1| = \dots = |R^p| = |C^1| = \dots = |C^p| = N$. Let $W \in \mathbb{Z}_{\geq 0}^{m \times n}$. Define*

$$R := \alpha_1 R^1 + \alpha_2 R^2 + \dots + \alpha_p R^p \quad \text{and} \quad C := \alpha_1 C^1 + \alpha_2 C^2 + \dots + \alpha_p C^p$$

and

$$T(R, C; W) := \sum_{A \in \mathbb{Z}^{m \times n} \cap \Pi(R, C)} \prod_{i=1}^m \prod_{j=1}^n w_{ij}^{a_{ij}}.$$

Then

$$\frac{\omega(N)T(R, C; W)}{\Omega(R)\Omega(C)} \geq \prod_{t=1}^p \left[\frac{T(R^t, C^t; W)}{\min\{\Omega(R^t), \Omega(C^t)\}} \right]^{\alpha_t}.$$

(See Definition 1 for the meaning of the functions ω , Ω .)

The matrix W may be thought of as attaching *weights* to the positions of an $m \times n$ matrix. If W is a 0–1 matrix, then $T(R, C; W)$ counts contingency tables with enforced zeroes in the positions (i, j) such that $w_{ij} = 0$. We easily prove Theorem 1 by recasting the margin and K -boundedness conditions as a regime of enforced zeroes (in a larger matrix).

2.1. Proof of Theorem 1. Assume the hypotheses of Theorem 1. Define vectors $\mathcal{R}^1, \mathcal{R}^2, \dots, \mathcal{R}^p, \mathcal{R} \in \mathbb{Z}_{\geq 0}^{m+n}$ and $\mathcal{C} \in \mathbb{Z}_{\geq 0}^{mn}$ by

$$\begin{aligned}\mathcal{R}^t &= (r_1^t, \dots, r_m^t, \tilde{c}_1^t, \dots, \tilde{c}_n^t), \\ \mathcal{R} &= (r_1, \dots, r_m, \tilde{c}_1, \dots, \tilde{c}_n), \\ \mathcal{C} &= (k_{11}, \dots, k_{1n}, k_{21}, \dots, k_{2n}, \dots, k_{m1}, \dots, k_{mn}).\end{aligned}$$

Observe that

$$|\mathcal{R}^1| = |\mathcal{R}^2| = \dots = |\mathcal{R}^p| = |\mathcal{C}| = |K|$$

and that

$$\mathcal{R} = \alpha_1 \mathcal{R}^1 + \alpha_2 \mathcal{R}^2 + \dots + \alpha_p \mathcal{R}^p.$$

Define $W = (w_{ij})$ as the $(m+n) \times (mn)$ matrix with

$$\begin{aligned}w_{i,(i-1)n+j} &= 1 && \text{for all } i = 1, \dots, m \text{ and } j = 1, \dots, n, \\ w_{m+j,(i-1)n+j} &= 1 && \text{for all } i = 1, \dots, m \text{ and } j = 1, \dots, n,\end{aligned}$$

and zeroes in all other positions.

Given a contingency table $A = (a_{ij}) \in \Pi_K(R, C)$, we may construct a table $\mathcal{A} = (a'_{ij}) \in \Pi(\mathcal{R}, \mathcal{C})$ by assigning

$$\begin{aligned}a'_{i,(i-1)n+j} &= a_{ij} && \text{for all } i = 1, \dots, m \text{ and } j = 1, \dots, n, \\ a'_{m+j,(i-1)n+j} &= k_{ij} - a_{ij} && \text{for all } i = 1, \dots, m \text{ and } j = 1, \dots, n,\end{aligned}$$

and zeroes in all other positions. This conversion is easily reversed, and thus gives a bijection between tables $A \in \Pi_K(R, C)$ and tables $\mathcal{A} \in \Pi(\mathcal{R}, \mathcal{C})$ which have enforced zeroes in all zero positions of W . That is,

$$T_K(R, C) = T(\mathcal{R}, \mathcal{C}; W). \quad (3)$$

Similarly,

$$T_K(R^t, C^t) = T(\mathcal{R}^t, \mathcal{C}; W) \quad (4)$$

for $t = 1, \dots, p$.

Substituting $\mathcal{R}^1, \dots, \mathcal{R}^p, \mathcal{R}$ for R^1, \dots, R^p, R in the statement of Theorem 5, as well as \mathcal{C} for C^1, \dots, C^p, C and $|K|$ for N , we obtain the conclusion

$$\frac{\omega(|K|)T(\mathcal{R}, \mathcal{C}; W)}{\Omega(\mathcal{R})\Omega(\mathcal{C})} \geq \prod_{t=1}^p \left[\frac{T(\mathcal{R}^t, \mathcal{C}; W)}{\min\{\Omega(\mathcal{R}^t), \Omega(\mathcal{C})\}} \right]^{\alpha_t}.$$

Using equations (3) and (4), and recalling the definitions of $\mathcal{R}, \mathcal{R}^t, \mathcal{C}$, we rewrite the above result as

$$\frac{\omega(|K|)T_K(R, C)}{\Omega(R)\Omega(\tilde{C})\Omega(K)} \geq \prod_{t=1}^p \left[\frac{T_K(R^t, C^t)}{\min\{\Omega(R^t)\Omega(\tilde{C}^t), \Omega(K)\}} \right]^{\alpha_t},$$

proving Theorem 1. ■

Remarks. The reader may notice that Theorem 1 can be stated in greater generality with only trivial modifications to the proof. For instance, $T_K(R, C)$ can be replaced by a weighted function $T_K(R, C; W)$, analogous to the function $T(R, C; W)$ in the statement of Theorem 1; also, given

$$K = \alpha_1 K^1 + \alpha_2 K^2 + \cdots + \alpha_p K^p,$$

Theorem 1 remains true when each instance of K on the right-hand side is replaced by K^t .

Theorem 5, here taken for granted, was originally derived from estimates for $T(R, C; W)$ obtained by Barvinok [2] via the theory of permanents and matrix scaling. Those estimates can be converted to the K -bounded setting in the manner illustrated above; however, we are able to bypass this step in the theory, as the more refined instrument of Theorem 5 is directly adaptable to our needs.

2.2. An honestly concave proxy for $\ln T_K(R, C)$. We now begin to assemble the ingredients for the proof of Theorem 2. Throughout this section, assume $K \in \mathbb{Z}_{>0}^{m \times n}$. We define a function which “smooths over” $\ln T_K(R, C)$:

Definition 8. For $R \in \mathbb{R}_{\geq 0}^m$, $C \in \mathbb{R}_{\geq 0}^n$, and $K \in \mathbb{Z}_{>0}^{m \times n}$, let

$$f(R, C) = f_K(R, C) := \max_{\substack{\alpha_1, \dots, \alpha_p \geq 0 \\ \alpha_1 + \dots + \alpha_p = 1 \\ \alpha_1 R^1 + \dots + \alpha_p R^p = R \\ \alpha_1 C^1 + \dots + \alpha_p C^p = C}} \sum_{t=1}^p \alpha_t \ln T_K(R^t, C^t).$$

(To be clear, the maximum is taken over choices of $p \geq 1$, $\alpha_1, \dots, \alpha_p$, R^1, \dots, R^p , and C^1, \dots, C^p which satisfy the indicated constraints, and for which the summation on the right is defined. If the maximum is taken over an empty set, then we regard it as $-\infty$.)

Note that the maximum in Definition 8 is well-defined (allowing for the $-\infty$ case), because there are finitely many pairs (R, C) for which $T_K(R, C) > 0$. It is redundant to allow any repetition among R^1, \dots, R^p or C^1, \dots, C^p , so the summation on the right takes on finitely many values.

Lemma 1 (Properties of $f(R, C)$).

- (i) $f(R, C) \geq \ln T_K(R, C)$.
- (ii) f is concave.

(iii) The domain of f (i.e., where $f > -\infty$) is a subset of $\Pi_K(R, C)$.

The proof of this lemma is straightforward, but the notation is cumbersome, so it is deferred to an appendix.

Lemma 2 (Quality of approximation). *Suppose $R \in \mathbb{Z}_{>0}^m$, $C \in \mathbb{Z}_{\geq 0}^n$, and $K \in \mathbb{Z}_{>0}^{m \times n}$. Define $\tilde{C} = (\tilde{c}_1, \dots, \tilde{c}_n)$ by $\tilde{c}_j := (\sum_{i=1}^m k_{ij}) - c_j$, and suppose that $\tilde{C} \in \mathbb{Z}_{>0}^n$. Then*

$$\begin{aligned} f_K(R, C) - \ln T_K(R, C) &\leq -\ln \sqrt{2\pi|K|} + \sum_{i=1}^m \ln \sqrt{2\pi r_i} + \sum_{j=1}^n \ln \sqrt{2\pi \tilde{c}_j} \\ &\quad + (m+n) \ln \left(\frac{e}{\sqrt{2\pi}} \right). \end{aligned}$$

Proof. By Stirling's formula,

$$n - \ln \sqrt{2\pi n} - \ln \left(\frac{e}{\sqrt{2\pi}} \right) \leq \ln \omega(n) \leq n - \ln \sqrt{2\pi n} \quad (5)$$

for $n \geq 1$.

Choose $\alpha_1, \dots, \alpha_p, R^1, \dots, R^p, C^1, \dots, C^p$ which achieve the maximum in Definition 8. Now apply Theorem 1 and (5):

$$\begin{aligned} f_K(R, C) - \ln T_K(R, C) &\leq \ln \left[\frac{\omega(|K|)}{\Omega(R)\Omega(\tilde{C})\Omega(K)} \cdot \prod_{t=1}^p \left(\min\{\Omega(R^t)\Omega(\tilde{C}^t), \Omega(K)\} \right)^{\alpha_t} \right] \\ &\leq \ln \left[\frac{\omega(|K|)}{\Omega(R)\Omega(\tilde{C})\Omega(K)} \cdot \prod_{t=1}^p \Omega(K)^{\alpha_t} \right] \\ &= \ln \frac{\omega(|K|)}{\Omega(R)\Omega(\tilde{C})} \\ &= \ln \omega(|K|) - \sum_{i=1}^m \ln \omega(r_i) - \sum_{j=1}^n \ln \omega(\tilde{c}_j) \\ &\leq |K| - \ln \sqrt{2\pi|K|} - \sum_i \left(r_i - \ln \sqrt{2\pi r_i} - \ln \left(\frac{e}{\sqrt{2\pi}} \right) \right) \\ &\quad - \sum_j \left(\tilde{c}_j - \ln \sqrt{2\pi \tilde{c}_j} - \ln \left(\frac{e}{\sqrt{2\pi}} \right) \right) \\ &\leq -\ln \sqrt{2\pi|K|} + \sum_{i=1}^m \ln \sqrt{2\pi r_i} + \sum_{j=1}^n \ln \sqrt{2\pi \tilde{c}_j} + (m+n) \ln \left(\frac{e}{\sqrt{2\pi}} \right). \quad \square \end{aligned}$$

2.3. Exact and approximate generating functions for tables. Theorem 2 refers to the following polynomial:

$$G(\mathbf{x}, \mathbf{y}) := \prod_{i=1}^m \prod_{j=1}^n [1 + x_i y_j + (x_i y_j)^2 + \cdots + (x_i y_j)^{k_{ij}}]$$

($\mathbf{x} = (x_1, \dots, x_m)$, $\mathbf{y} = (y_1, \dots, y_n)$).

Lemma 3. *G is a generating function for K -bounded contingency tables; that is,*

$$G(\mathbf{x}, \mathbf{y}) = \sum_R \sum_C T_K(R, C) \mathbf{x}^R \mathbf{y}^C, \quad (6)$$

where the sum is taken over all possible margins R, C (of lengths m and n).

Proof. Trivial. \square

In principle, we can “compute” $T_K(R, C)$ by expanding $G(\mathbf{x}, \mathbf{y})$ and extracting the coefficient of $\mathbf{x}^R \mathbf{y}^C$. This is of course not practical, but we might estimate this coefficient by

$$\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C};$$

indeed, this is an *upper* bound on $T_K(R, C)$, as may be readily seen by dividing both sides of (6) by $\mathbf{x}^R \mathbf{y}^C$. To bound $T_K(R, C)$ from the other side, we replace $G(\mathbf{x}, \mathbf{y})$ by an approximate version with smoother coefficients:

Definition 9. *Let*

$$\tilde{G}(\mathbf{x}, \mathbf{y}) := \sum_R \sum_C e^{f(R, C)} \mathbf{x}^R \mathbf{y}^C,$$

where the sum is taken over all integer margins (R, C) such that $f(R, C) > -\infty$.

(See Definition 8 for the meaning of $f(R, C)$.)

We will find the following lemma useful, as it will allow us to pick out any nonzero term of $\tilde{G}(\mathbf{x}, \mathbf{y})$ as the largest:

Lemma 4. *For any (R_*, C_*) in the relative interior of the domain of f , there exist $\mathbf{x}_*, \mathbf{y}_* > 0$ such that the function*

$$\Phi(R, C) := e^{f(R, C)} \mathbf{x}_*^R \mathbf{y}_*^C$$

attains its maximum at $R = R_, C = C_*$.*

Proof. Recall that f is concave; therefore, its graph has a supporting hyperplane over (R_*, C_*) . Let such a hyperplane have outward-pointing normal vector $(u_1, \dots, u_m, v_1, \dots, v_n, 1)$. Set

$$\mathbf{x}_* = (x_1, \dots, x_m) = (e^{-u_1}, \dots, e^{-u_m}) \quad \text{and} \quad \mathbf{y}_* = (y_1, \dots, y_n) = (e^{-v_1}, \dots, e^{-v_n}).$$

Then

$$\phi(R, C) := f(R, C) + \sum_{i=1}^m r_i \ln x_i + \sum_{j=1}^n c_j \ln y_j$$

is concave with respect to R and C , and attains a critical point (hence its global maximum) at (R_*, C_*) . Therefore, so does $\Phi(R, C) = e^{\phi(R, C)}$. \square

2.4. Proof of Theorem 2. Assume the hypotheses of Theorem 2. Using Lemma 4, choose $\mathbf{x}_*, \mathbf{y}_*$ so that $e^{f(R, C)} \mathbf{x}_*^R \mathbf{y}_*^C$ is the largest term in the expansion of $\tilde{G}(\mathbf{x}, \mathbf{y})$, evaluated at $\mathbf{x} = \mathbf{x}_*$ and $\mathbf{y} = \mathbf{y}_*$. Thus

$$\frac{\tilde{G}(\mathbf{x}_*, \mathbf{y}_*)}{\mathbf{x}_*^R \mathbf{y}_*^C} \leq [\# \text{ of terms of } \tilde{G} \text{ with nonzero coeffs.}] \cdot e^{f(R, C)}.$$

The number of terms of \tilde{G} is at most

$$\mathcal{N} := \prod_{i=1}^m \left(1 + \sum_{j=1}^n k_{ij} \right) + \prod_{j=1}^n \left(1 + \sum_{i=1}^m k_{ij} \right),$$

since $T_K(R, C) > 0$ implies that R and C do not exceed the margins of K .

Let the symbol \heartsuit denote the quantity

$$-\ln \sqrt{2\pi|K|} + \sum_{i=1}^m \ln \sqrt{2\pi r_i} + \sum_{j=1}^n \ln \sqrt{2\pi \tilde{c}_j} + (m+n) \ln \left(\frac{e}{\sqrt{2\pi}} \right),$$

last seen in Lemma 2.

We deduce the following chain of inequalities:

$$\begin{aligned} \ln \left(\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right) &\geq \ln T_K(R, C) \\ &\geq f(R, C) - \heartsuit \\ &\geq \ln \left(\frac{\tilde{G}(\mathbf{x}_*, \mathbf{y}_*)}{\mathbf{x}_*^R \mathbf{y}_*^C \cdot \mathcal{N}} \right) - \heartsuit \\ &\geq \ln \left(\inf_{x_i, y_j > 0} \frac{\tilde{G}(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C \cdot \mathcal{N}} \right) - \heartsuit \\ &\geq \ln \left(\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right) - \ln \mathcal{N} - \heartsuit. \end{aligned} \tag{7}$$

Now we consider the cloning of the margins. Let $G^{(s)}$ denote the generating function for $K^{(s)}$ -bounded contingency tables. Letting

$$\begin{aligned} \mathbf{x}^{(s)} &:= (\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^s) \\ &= (x_1^1, \dots, x_m^1, x_1^2, \dots, x_m^2, \dots, x_1^s, \dots, x_m^s), \end{aligned}$$

and defining $\mathbf{y}^{(s)}$ similarly, we note that

$$\frac{G^{(s)}(\mathbf{x}^{(s)}, \mathbf{y}^{(s)})}{[\mathbf{x}^{(s)}]^{R^{(s)}} [\mathbf{y}^{(s)}]^{C^{(s)}}} = \prod_{k=1}^s \prod_{\ell=1}^s \frac{G(\mathbf{x}^k, \mathbf{y}^\ell)}{(\mathbf{x}^k)^R (\mathbf{y}^\ell)^C}.$$

From this it follows that

$$\frac{1}{s^2} \ln \left(\inf_{x_i^k, y_j^\ell > 0} \frac{G^{(s)}(\mathbf{x}^{(s)}, \mathbf{y}^{(s)})}{[\mathbf{x}^{(s)}]^{R^{(s)}} [\mathbf{y}^{(s)}]^{C^{(s)}}} \right) = \ln \left(\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right)$$

for all $s \geq 1$.

Inspection of the formulas for $\ln \mathcal{N}$ and \heartsuit shows that both of these terms from (7) have growth of order $O(s \ln s)$ as $s \rightarrow \infty$. Therefore, by (7),

$$\frac{1}{s^2} \ln T_{K^{(s)}}(R^{(s)}, C^{(s)}) = \ln \left(\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right) + O\left(\frac{\ln s}{s}\right),$$

from which Theorem 2 follows. ■

3. ENTROPY-BASED ESTIMATES FOR $T_K(R, C)$

In section 1.3, we introduced the functions $H_\kappa^{\max}(x)$ and $p(x; \kappa), q(x; \kappa)$; refer to equations (1), (2) for an implicit description of the latter. We now list a few useful facts about H_κ^{\max} :

Lemma 5. *Let $p = p(x; \kappa)$, $q = q(x; \kappa)$.*

- (i) H_κ^{\max} is strictly concave on its domain.
- (ii) $H_\kappa^{\max}(x) = -[\ln p + x \ln q]$.
- (iii) For $0 < x < \kappa$, $\frac{d}{dx} H_\kappa^{\max}(x) = -\ln q$.

Proof. First we prove claim (i). Let $x, y \in [0, \kappa]$ and $\alpha, \beta > 0$ such that $\alpha + \beta = 1$. We wish to prove that

$$H_\kappa^{\max}(\alpha x + \beta y) > \alpha H_\kappa^{\max}(x) + \beta H_\kappa^{\max}(y).$$

Let X and Y be independent random variables with distributions $TG(x; \kappa)$ and $TG(y; \kappa)$, respectively. Define a random variable Z whose distribution is a *mixture* of X and Y with weights α and β ; that is,

$$\Pr[Z = t] = \alpha p(x; \kappa) q(x; \kappa)^t + \beta p(y; \kappa) q(y; \kappa)^t \quad \text{for } t = 0, 1, \dots, \kappa.$$

Then

$$\mathbf{E}[Z] = \alpha x + \beta y$$

and

$$H[Z] > \alpha H[X] + \beta H[Y]$$

(where H denotes the entropy, which is well-known to be strictly concave with respect to mixture). But

$$H_{\kappa}^{\max}(\alpha x + \beta y) \geq H[Z],$$

since $H_{\kappa}^{\max}(\alpha x + \beta y)$ is the maximum entropy achieved by any random variable supported on $\{0, 1, \dots, \kappa\}$ with expectation $\alpha x + \beta y$. This concludes the proof of (i).

Claim (ii) follows readily from equations (1), (2) and the properties of logarithms. Let $p = p(x; \kappa)$, $q = q(x; \kappa)$. Then by definition of entropy, we have

$$\begin{aligned} H_{\kappa}^{\max}(x) &= -[p \ln p + pq \ln(pq) + pq^2 \ln(pq^2) + \dots + pq^{\kappa} \ln(pq^{\kappa})] \\ &= -[p \ln p + pq(\ln p + \ln q) + pq^2(\ln p + 2 \ln q) + \dots + pq^{\kappa}(\ln p + \kappa \ln q)] \\ &= -[(p + pq + pq^2 + \dots + pq^{\kappa})(\ln p) + (pq + 2pq^2 + \dots + \kappa pq^{\kappa})(\ln q)] \\ &= -[\ln p + x \ln q]. \end{aligned}$$

Differentiating this formula with respect to x , and again using equations (1) and (2), we obtain

$$\begin{aligned} (H_{\kappa}^{\max})'(x) &= -\frac{p'}{p} - x \cdot \frac{q'}{q} - \ln q \\ &= p \cdot \left(\frac{1}{p}\right)' - p(q + 2q^2 + \dots + \kappa q^{\kappa}) \cdot \frac{q'}{q} - \ln q \\ &= p \cdot \left(\frac{1}{p}\right)' - pq'(1 + 2q + \dots + \kappa q^{\kappa-1}) - \ln q \\ &= p \cdot \left(\frac{1}{p}\right)' - p \cdot \left(\frac{1}{p}\right)' - \ln q \\ &= -\ln q. \end{aligned}$$

This proves claim (iii). \square

Like the “ $(\kappa + 1)$ -nomial coefficients” of section 1.4, the functions H_{κ}^{\max} , p , and q admit simple explicit formulas only when $\kappa = 1$ or $\kappa = \infty$. To wit:

$$\begin{aligned} H_1^{\max}(x) &= -x \ln x - (1 - x) \ln(1 - x) & p(x; 1) &= 1 - x & q(x; 1) &= \frac{x}{1 - x} \\ H_{\infty}^{\max}(x) &= (x + 1) \ln(x + 1) - x \ln x & p(x; \infty) &= \frac{1}{x + 1} & q(x; \infty) &= \frac{x}{x + 1} \end{aligned}$$

In fact, there is a close relationship between these functions and the $(\kappa + 1)$ -nomial coefficients:

Lemma 6. Let $\kappa \in \mathbb{Z}_{>0} \cup \{\infty\}$. Let n, r be integers ($n > 0$, $0 \leq r \leq n\kappa$). Then

$$\lim_{s \rightarrow \infty} \frac{1}{s} \ln \binom{sn}{sr}_{\kappa} = nH_{\kappa}^{\max} \left(\frac{r}{n} \right).$$

Proof. Let X_1, X_2, \dots be independent random variables, each with distribution $TG\left(\frac{r}{n}; \kappa\right)$. Let $X = (X_1, \dots, X_{sn})$.

Observe that if $\mathbf{x}, \mathbf{x}' \in \{0, 1, \dots, \kappa\}^s$, then

$$\frac{\Pr[X = \mathbf{x}']}{\Pr[X = \mathbf{x}]} = q^{|\mathbf{x}'| - |\mathbf{x}|}$$

(where $|\mathbf{x}| := \sum_{i=1}^s x_i$). In particular, all values of X with equal sum of coordinates are equiprobable. Let \mathbf{x}_* denote an arbitrary value for X satisfying $|\mathbf{x}_*| = sr$.

Recall that the *Shannon self-information* of a value $X = \mathbf{x}$ is defined as

$$I(\mathbf{x}) := -\ln \Pr[X = \mathbf{x}];$$

the entropy of X is the expected self-information of its value. Thus we have

$$\begin{aligned} snH_{\kappa}^{\max} \left(\frac{r}{n} \right) &= H[X] = \mathbf{E}[I(X)] \\ &= I(\mathbf{x}_*) - (\ln q) \mathbf{E}[|X| - sr] \\ &= I(\mathbf{x}_*) \\ &= -\ln \left[\binom{sn}{sr}_{\kappa}^{-1} \cdot \Pr[|X| = sr] \right] \\ &= \ln \binom{sn}{sr}_{\kappa} - \ln \Pr[|X| = sr]. \end{aligned} \quad (8)$$

Note that the probability mass function for each X_i is log-concave on \mathbb{Z} . We apply a local limit theorem of Bender (see Appendix, Theorem 7) using

$$\zeta_p = X_1 + \dots + X_p, \quad \sigma_p^2 = p \cdot \text{Var}(X_1), \quad \mu_p = p \cdot \frac{r}{n}, \quad \text{and} \quad x = 0,$$

with the normality hypothesis secured via Lyapunov's central limit theorem (Appendix, Theorem 6), to infer

$$\lim_{p \rightarrow \infty} \sigma_p \Pr \left[\zeta_p = \left\lfloor p \cdot \frac{r}{n} \right\rfloor \right] = \frac{1}{\sqrt{2\pi}},$$

and thence

$$\Pr[|X| = sr] \sim (2\pi sn \text{Var}(X_1))^{-1/2} = \Theta(s^{-1/2}).$$

Substituting into (8), we conclude that

$$snH_{\kappa}^{\max} \left(\frac{r}{n} \right) = \ln \binom{sn}{sr}_{\kappa} - \Theta(\ln s);$$

in particular, this proves the lemma. \square

3.1. A dual to the optimization problem in Theorem 2. In Theorem 2, we “computed” $\ln T_K(R, C)$ asymptotically as

$$\ln \left(\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right). \quad (9)$$

We now show:

Lemma 7. *Suppose $k_{ij} = \kappa$ for $1 \leq i \leq m$, $1 \leq j \leq n$. Then the quantity given in (9) is equal to*

$$\max_{Z \in \Pi_\kappa(R, C)} \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max}(z_{ij}). \quad (10)$$

Proof. By Lemma 5, $H_\kappa^{\max}(x)$ is strictly concave. Also, $(H_\kappa^{\max})'(x) = \ln q(x; \kappa)$ approaches ∞ as $x \rightarrow 0$ and $-\infty$ as $x \rightarrow \kappa$; therefore, the maximum in (10) is well-defined and is attained in the relative interior of $\Pi_\kappa(R, C)$. For the remainder of this proof, let Z denote the (unique) location at which the maximum is attained, and let $p_{ij} := p(z_{ij}; \kappa)$, $q_{ij} := q(z_{ij}; \kappa)$.

Since Z is in the interior of $\Pi_\kappa(R, C)$, the local defining equations for $\Pi_\kappa(R, C)$ at Z are just

$$\sum_{j=1}^n a_{ij} = r_i \quad (1 \leq i \leq m) \quad \text{and} \quad \sum_{i=1}^m a_{ij} = c_j \quad (1 \leq j \leq n).$$

Introducing Lagrange multipliers for these constraints, we infer that $\ln q_{ij} = \lambda_i + \mu_j$ for some constants $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_n$. Define $\xi_i := e^{\lambda_i}$, $\eta_j = e^{\mu_j}$; thus $q_{ij} = \xi_i \eta_j$. Dividing equation (2) by equation (1) (see section 1.3), we obtain

$$z_{ij} = \frac{\xi_i \eta_j + 2(\xi_i \eta_j)^2 + \dots + \kappa(\xi_i \eta_j)^\kappa}{1 + \xi_i \eta_j + (\xi_i \eta_j)^2 + \dots + (\xi_i \eta_j)^\kappa}.$$

For real-valued $\mathbf{t} = (t_1, \dots, t_m)$ and $\mathbf{s} = (s_1, \dots, s_n)$, let

$$\begin{aligned} \psi(\mathbf{t}, \mathbf{s}) &:= \ln \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \Big|_{\substack{x_i = e^{t_i} \\ y_j = e^{s_j}}} \\ &= - \sum_{i=1}^m r_i t_i - \sum_{j=1}^n c_j s_j + \sum_{i=1}^m \sum_{j=1}^n \ln \left(1 + e^{t_i + s_j} + e^{2(t_i + s_j)} + \dots + e^{\kappa(t_i + s_j)} \right). \end{aligned}$$

This function is strictly convex, and has a critical point (hence a global minimum) at (\mathbf{t}, \mathbf{s}) if and only if the gradient is zero, that is, if

$$\begin{aligned} r_i &= \sum_{j=1}^n \frac{e^{t_i + s_j} + 2e^{2(t_i + s_j)} + \dots + \kappa e^{\kappa(t_i + s_j)}}{1 + e^{t_i + s_j} + e^{2(t_i + s_j)} + \dots + e^{\kappa(t_i + s_j)}}, & 1 \leq i \leq m \\ \text{and } c_j &= \sum_{i=1}^m \frac{e^{t_i + s_j} + 2e^{2(t_i + s_j)} + \dots + \kappa e^{\kappa(t_i + s_j)}}{1 + e^{t_i + s_j} + e^{2(t_i + s_j)} + \dots + e^{\kappa(t_i + s_j)}}, & 1 \leq j \leq n. \end{aligned}$$

These conditions are satisfied at $\mathbf{t} = (\lambda_1, \dots, \lambda_m)$ and $\mathbf{s} = (\mu_1, \dots, \mu_n)$. The minimum value of ψ is thus

$$\begin{aligned}
\psi(\mathbf{t}, \mathbf{s}) &= -\sum_{i=1}^m r_i \lambda_i - \sum_{j=1}^n c_j \mu_j + \sum_{i=1}^m \sum_{j=1}^n \ln(1 + \xi_i \eta_j + (\xi_i \eta_j)^2 + \dots + (\xi_i \eta_j)^\kappa) \\
&= \sum_{i=1}^m \sum_{j=1}^n [-z_{ij}(\lambda_i + \mu_j) + \ln(1 + q_{ij} + q_{ij}^2 + \dots + q_{ij}^\kappa)] \\
&= \sum_{i=1}^m \sum_{j=1}^n \left[-z_{ij} \ln q_{ij} + \ln \left(\frac{1}{p_{ij}} \right) \right] \\
&= \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max}(z_{ij}).
\end{aligned}$$

This proves the lemma. \square

3.2. Proof of Theorem 3. Part (i) follows directly from Theorem 2 and Lemma 7.

For part (ii), recall that

$$I_\kappa(R, C) = \binom{mn}{N}_\kappa^{-1} \prod_{i=1}^m \binom{n}{r_i}_\kappa \prod_{j=1}^n \binom{m}{c_j}_\kappa$$

(Definition 7). Thus

$$I_\kappa(R^{(s)}, C^{(s)}) = \binom{s^2 mn}{s^2 N}_\kappa^{-1} \left(\prod_{i=1}^m \binom{sn}{sr_i}_\kappa \right)^s \left(\prod_{j=1}^n \binom{sm}{sc_j}_\kappa \right)^s.$$

Applying Lemma 6, we obtain

$$\begin{aligned}
\ln I_\kappa(R^{(s)}, C^{(s)}) &= -\left[s^2 mn H_\kappa^{\max} \left(\frac{N}{mn} \right) + o(s^2) \right] + s \sum_{i=1}^m \left[sn H_\kappa^{\max} \left(\frac{r_i}{n} \right) + o(s) \right] \\
&\quad + s \sum_{j=1}^n \left[sm H_\kappa^{\max} \left(\frac{c_j}{m} \right) + o(s) \right] \\
&= s^2 \left[-mn H_\kappa^{\max} \left(\frac{N}{mn} \right) + n \sum_{i=1}^m H_\kappa^{\max} \left(\frac{r_i}{n} \right) + m \sum_{j=1}^n H_\kappa^{\max} \left(\frac{c_j}{m} \right) + o(1) \right].
\end{aligned}$$

This completes the proof of Theorem 3. \blacksquare

3.3. The entropy loss function. The following function plays a key role in the proof of Theorem 4:

Definition 10. Fix $\kappa \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$. Given nonnegative $\alpha_1, \alpha_2, \dots, \alpha_n$ such that $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$, let

$$J(r) = J_{\alpha, \kappa}(r) := nH_{\kappa}^{\max}\left(\frac{r}{n}\right) - \sum_{j=1}^n H_{\kappa}^{\max}(r\alpha_j)$$

for all $r \geq 0$ such that $r\alpha_1, r\alpha_2, \dots, r\alpha_n \leq \kappa$.

In the spirit of the remarks at the beginning of section 1.3, we propose the following (informal) interpretation for $J(r)$. (The reader who is interested only in formal proof may skip to section 3.4.)

Suppose A is an $m \times n$ contingency table, about which we know only that the sum of entries is N . In order to guess what A looks like, we might sample mn entries independently from $TG\left(\frac{N}{mn}; \kappa\right)$. The resulting matrix might not have sum of entries exactly equal to N , but at least it is correct in expectation, and all actual tables with sum of entries N are equally likely to be chosen. The entropy of our random model is $mnH_{\kappa}^{\max}\left(\frac{N}{mn}\right)$.

If we subsequently learn that A has column sums $C = (c_1, \dots, c_n)$, then we might revise our model by drawing entries in the j^{th} column from $TG\left(\frac{c_j}{m}; \kappa\right)$. The entropy of the random model is then $m \sum_{j=1}^n H_{\kappa}^{\max}\left(\frac{c_j}{m}\right)$. The information gained may be measured by the entropy lost, which is equal to $mJ\left(\frac{N}{m}\right)$ for $\alpha_j = \frac{c_j}{N}$ ($j = 1, \dots, n$).

The same comparison may be made in the presence of known row sums $R = (r_1, \dots, r_m)$. Before learning C , we sample entry (i, j) from $TG\left(\frac{r_i}{n}; \kappa\right)$; after learning C , we sample entry (i, j) from $TG\left(\frac{r_i c_j}{N}; \kappa\right)$. Of course, this guess is entirely naïve—it essentially regards the rank 1 table as typical, which is a doubtful assumption. We do not propose that this is the *best* guess, but it does have the advantage of being computable. The entropy lost to this model when incorporating C in the presence of known R is

$$\sum_{i=1}^m J(r_i),$$

where $\alpha_j = \frac{c_j}{N}$ ($j = 1, \dots, n$). Note that this quantity is defined if and only if the rank 1 table has all entries $\leq \kappa$.

We would like to know if the row margin R and the column margin C are positively correlated. Intuitively, this means that the revelation of C produces less *surprise* (entropy loss) when R is known in advance than when R is not known. In our random model, this is expressed by the inequality

$$mJ\left(\frac{N}{m}\right) \geq J(r_1) + J(r_2) + \dots + J(r_m). \quad (11)$$

Now let us see if we can make a formal argument out of the intuition.

Theorem 3, part (i), implies that

$$\lim_{s \rightarrow \infty} \frac{1}{s^2} \ln T_\kappa(R^{(s)}, C^{(s)}) \geq \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max} \left(\frac{r_i c_j}{N} \right);$$

again, the rank 1 matrix is chosen here purely out of convenience (we could have substituted any $Z \in \Pi_\kappa(R, C)$). Together with part (ii), this means that in order to prove

$$\lim_{s \rightarrow \infty} \frac{1}{s^2} \ln T_\kappa(R^{(s)}, C^{(s)}) \geq \lim_{s \rightarrow \infty} \frac{1}{s^2} \ln I_\kappa(R^{(s)}, C^{(s)}),$$

it suffices to show that

$$mn H_\kappa^{\max} \left(\frac{N}{mn} \right) \geq - \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max} \left(\frac{r_i c_j}{N} \right) + m \sum_{j=1}^n H_\kappa^{\max} \left(\frac{c_j}{m} \right) + n \sum_{i=1}^m H_\kappa^{\max} \left(\frac{r_i}{n} \right)$$

or, equivalently (by rearranging terms),

$$mn H_\kappa^{\max} \left(\frac{N}{mn} \right) - m \sum_{j=1}^n H_\kappa^{\max} \left(\frac{c_j}{m} \right) \geq n \sum_{i=1}^m H_\kappa^{\max} \left(\frac{r_i}{n} \right) - \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max} \left(\frac{r_i c_j}{N} \right).$$

This is precisely what is asserted by the inequality (11). In order to prove Theorem 4, we will show that under its hypotheses, (11) is true (and meaningful, i.e. A^{ind} has entries $\leq \kappa$).

3.4. Proof of Theorem 4. Assume the hypotheses of Theorem 4; note in particular that $\kappa \geq 2$. Let

$$\alpha_j := \frac{c_j}{N} \quad (1 \leq j \leq n).$$

Consider the function

$$\phi(x) := x^2 (H_\kappa^{\max})''(x) = -x^2 \cdot \frac{q'(x; \kappa)}{q(x; \kappa)}$$

(all derivatives being with respect to x ; the second equality follows from Lemma 5(iii)).

The above formula defines $\phi(x)$ only for $0 \leq x \leq \kappa$, but we claim that $\phi(x)$ can be extended analytically to a neighborhood of $x = 0$.

Proof of claim: Equations (1) and (2) (section 1.3) yield

$$x = \frac{q + 2q^2 + \cdots + \kappa q^\kappa}{1 + q + q^2 + \cdots + q^\kappa},$$

where $q = q(x; \kappa)$. Although this formula has only been assigned meaning for $q \geq 0$, it shows that x (as a function of q) can be extended analytically to a neighborhood of $q = 0$; the Maclaurin series is $x = q + q^2 + O(q^3)$. Since $\frac{dx}{dq} \neq 0$ at $q = 0$, it follows that the inverse function $q(x; \kappa)$ is also defined and analytic in a neighborhood of $x = 0$, with Maclaurin series $q = x - x^2 + O(x^3)$. Applying l'Hôpital's rule, we see

that the singularity of ϕ at $x = 0$ is removable, so $\phi(x)$ is locally analytic there. \square

We compute the Maclaurin series of $\phi(x)$:

$$\phi(x) = -x \cdot \frac{1 - 2x + O(x^2)}{1 - x + O(x^2)} = -x + x^2 + O(x^3).$$

Since the coefficient of x^2 is positive, $\phi(x)$ is strictly convex in a neighborhood of $x = 0$. Choose $\delta \in (0, 1)$ such that $\phi(x)$ is strictly convex in the interval $|x| \leq \delta\kappa$.

Because $\delta < 1$, $J(r)$ (see definition 10) is defined and differentiable at $r = r_1, \dots, r_m$. Differentiating, we have

$$J'(r) = (H_\kappa^{\max})' \left(\frac{r}{n} \right) - \sum_{j=1}^n \alpha_j (H_\kappa^{\max})'(r\alpha_j)$$

and

$$\begin{aligned} J''(r) &= \frac{1}{n} (H_\kappa^{\max})'' \left(\frac{r}{n} \right) - \sum_{j=1}^n \alpha_j^2 (H_\kappa^{\max})''(r\alpha_j) \\ &= \frac{n}{r^2} \phi \left(\frac{r}{n} \right) - \sum_{j=1}^n \frac{1}{r^2} \phi(r\alpha_j). \end{aligned}$$

By the convexity of $\phi(x)$, we have $J''(r) \leq 0$ for $0 < r \leq \frac{\delta\kappa}{\max\{\alpha_1, \dots, \alpha_n\}}$; the inequality is strict if $\alpha_1, \dots, \alpha_n$ are not all equal. Therefore, $J(r)$ is concave on (the closure of) that interval, and strictly concave if $\alpha_1, \dots, \alpha_n$ are not all equal. By our assumption that $r_i c_j \leq \delta\kappa N$, it follows that r_1, \dots, r_m are in that interval.

Thus, inequality (11) holds:

$$mJ\left(\frac{N}{m}\right) \geq J(r_1) + J(r_2) + \dots + J(r_m),$$

with strict inequality if $\alpha_1, \dots, \alpha_n$ are not all equal and r_1, \dots, r_m are also not all equal.

When the function J is evaluated throughout this inequality, we obtain

$$mnH_\kappa^{\max}\left(\frac{N}{mn}\right) - m \sum_{j=1}^n H_\kappa^{\max}\left(\frac{c_j}{m}\right) \geq n \sum_{i=1}^m H_\kappa^{\max}\left(\frac{r_i}{n}\right) - \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max}\left(\frac{r_i c_j}{N}\right).$$

As explained at the end of section 3.3, this implies the last link, and Theorem 3 implies the first, in this chain of inequalities:

$$\begin{aligned}
\lim_{s \rightarrow \infty} \frac{1}{s^2} \ln T_\kappa(R^{(s)}, C^{(s)}) &\geq \max_{Z \in \Pi_\kappa(R, C)} \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max}(z_{ij}) \\
&\geq \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max}\left(\frac{r_i c_j}{N}\right) \\
&\geq \lim_{s \rightarrow \infty} \frac{1}{s^2} \ln I_\kappa(R^{(s)}, C^{(s)}).
\end{aligned}$$

If $\alpha_1, \dots, \alpha_n$ are not all equal and r_1, \dots, r_m are not all equal, then the last inequality in this chain is strict. This completes the proof of Theorem 4. ■

3.5. Prospects for negative correlation of margins. Recall that for $\kappa = 1$, all pairs of margins (R, C) have either zero or negative asymptotic correlation (specifically, negative unless either R or C is a constant vector). For $\kappa = \infty$, the sign of correlation is reversed. We expect that these are the only “pure” cases: that is, when $1 < \kappa < \infty$, there are some positively correlated pairs of margins as well as some negatively correlated pairs. Theorem 4 asserts half of this conjecture: for $\kappa \geq 2$, any sufficiently sparse margins are asymptotically positively correlated. (By symmetry, “co-sparse margins”—those which force most entries to be close to κ —are also positively correlated.)

Numerical evidence and heuristic arguments suggest that, for all $\kappa < \infty$, margins which are neither sparse nor co-sparse—or more specifically, close to $R = (\frac{n\kappa}{2}, \dots, \frac{n\kappa}{2})$ and $C = (\frac{m\kappa}{2}, \dots, \frac{m\kappa}{2})$ —are negatively correlated. For example, we have used Theorem 3 to compute

$$\lim_{s \rightarrow \infty} \frac{1}{s^2} [\ln T_\kappa(R^{(s)}, C^{(s)}) - \ln I_\kappa(R^{(s)}, C^{(s)})] \quad (12)$$

for margins of the form $R = C = (\gamma, \gamma + \varepsilon, \gamma + 2\varepsilon, \dots, \gamma + (n-1)\varepsilon)$, $n = 2, 3, 4, 5$. In these tests, when $\kappa = 2$ and $\varepsilon = .02$, expression (12) turns out to be negative (indicating negative correlation of the margins) roughly when $.09n < \gamma < 1.89n$, suggesting a threshold of $\delta \approx .05$. When $\kappa = 10$ and $\varepsilon = .1$, expression (12) is negative roughly when $1.5n < \gamma < 8.5n$, suggesting $\delta \approx .15$. As $\kappa \rightarrow \infty$, the sharp value of δ in Theorem 4 appears to grow, but a threshold remains.

An intuitive gloss on this phenomenon is that the distribution $TG(x; \kappa)$ “looks like” a geometric distribution when $x \approx 0$ (or $x \approx \kappa$), but looks more like a Bernoulli distribution when x is at neither extreme. In the former case, the “lid” κ (or the floor 0) is remote from typical values, so the behavior observed when $\kappa = \infty$ dominates. In the latter case, the $\kappa = 1$ behavior seems to dominate.

The fundamental difference between these cases is suggested by the function $\phi(x)$ in the proof of Theorem 4. When $\kappa = \infty$, this function is convex throughout its

domain; when $\kappa = 1$, it is concave; and when $1 < \kappa < \infty$, this function is convex near the origin, but has an inflection point.

We can show that $\phi(x)$ is concave for $x \approx \frac{\kappa}{2}$, so what are the obstacles to a reverse Theorem 4? There are two. In the proof of Theorem 4, we relied on the fact that

$$\max_{Z \in \Pi_\kappa(R, C)} \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max}(z_{ij}) \geq \sum_{i=1}^m \sum_{j=1}^n H_\kappa^{\max}\left(\frac{r_i c_j}{N}\right), \quad (13)$$

allowing us to use the rank 1 matrix $A^{\text{ind}} = \left(\frac{r_i c_j}{N}\right)$ as a proxy for the unknown Z which achieves the maximum. This matrix does not necessarily have entries $\leq \kappa$; we were able to assume that it does only because our assumption of sparse margins did double duty. This is the first obstacle to a reverse Theorem 4; the second is that, even if we could find another plausible candidate for Z , we could not make an assumption like (13) with the inequality reversed. Thus, to prove a negative correlation result, we believe it is crucial to understand something about *where* the maximum on the left-hand side of (13) is achieved.

4. APPENDIX

This section contains some matter which was deferred from earlier sections.

4.1. Proof of Lemma 1. Claim (i) is trivial, since we can set $p = 1$, $\alpha_1 = 1$, $R^1 = R$, $C^1 = C$ in Definition 8.

For claim (ii), it suffices to show that if $\alpha + \beta = 1$, then

$$\alpha f(R^1, C^1) + \beta f(R^2, C^2) \leq f(\alpha R^1 + \beta R^2, \alpha C^1 + \beta C^2). \quad (14)$$

By Definition 8, there exist $\gamma_1, \dots, \gamma_p \geq 0$; R^{11}, \dots, R^{1p} ; and C^{11}, \dots, C^{1p} such that

$$\begin{aligned} \sum_{t=1}^p \gamma_t &= 1, & \sum_{t=1}^p \gamma_t R^{1t} &= R, & \sum_{t=1}^p \gamma_t C^{1t} &= C, \\ \text{and } f(R^1, C^1) &= \sum_{t=1}^p \gamma_t \ln T_K(R^{1t}, C^{1t}). \end{aligned}$$

Likewise, there exist $\delta_1, \dots, \delta_q \geq 0$; R^{21}, \dots, R^{2q} ; and C^{21}, \dots, C^{2q} such that

$$\begin{aligned} \sum_{t=1}^q \delta_t &= 1, & \sum_{t=1}^q \delta_t R^{2t} &= R, & \sum_{t=1}^q \delta_t C^{2t} &= C, \\ \text{and } f(R^2, C^2) &= \sum_{t=1}^q \delta_t \ln T_K(R^{2t}, C^{2t}). \end{aligned}$$

Note that

$$\sum_{t=1}^p \alpha \gamma_t + \sum_{t=1}^q \beta \delta_t = 1, \quad \sum_{t=1}^p \alpha \gamma_t R^{1t} + \sum_{t=1}^q \beta \delta_t R^{2t} = \alpha R^1 + \beta R^2,$$

$$\text{and} \quad \sum_{t=1}^p \alpha \gamma_t C^{1t} + \sum_{t=1}^q \beta \delta_t C^{2t} = \alpha C^1 + \beta C^2;$$

applying Definition 8 to $f(\alpha R^1 + \beta R^2, \alpha C^1 + \beta C^2)$, we obtain equation (14) and thus claim (ii).

It is clear that f is defined only on the convex hull of all (R, C) for which $T_K(R, C) > 0$; this region is a subset of $\Pi_K(R, C)$, proving claim (iii). \square

4.2. Limit theorems of Lyapunov and Bender. We make use of the following theorems in the proof of Lemma 6:

Theorem 6 (Lyapunov's central limit theorem). *Suppose that $(X_n : n \in \mathbb{N})$ is a sequence of independent random variables, such that $\mu_n := \mathbf{E}[X_n]$ and $\sigma_n^2 := \text{Var}[X_n]$ are finite. Let $\zeta_n = X_1 + \cdots + X_n$, and define $m_n := \mathbf{E}[\zeta_n] = \mu_1 + \cdots + \mu_n$, $s_n^2 := \text{Var}[\zeta_n] = \sigma_1^2 + \cdots + \sigma_n^2$. If*

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathbf{E}[|X_k - \mu_k|^{2+\delta}] = 0$$

for some $\delta > 0$, then

$$\lim_{n \rightarrow \infty} \mathbf{Pr}[\zeta_n < s_n x + m_n] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

for all $x \in \mathbb{R}$.

Theorem 7 (Bender local limit theorem). *Suppose that $(\zeta_n : n \in \mathbb{N})$ is a sequence of integer-valued random variables and (σ_n) and (μ_n) are sequences of real numbers, such that*

$$\lim_{n \rightarrow \infty} \mathbf{Pr}[\zeta_n < \sigma_n x + \mu_n] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

for all $x \in \mathbb{R}$. Also suppose that $\sigma_n \rightarrow \infty$ as $n \rightarrow \infty$. Further, suppose that, for every n , the sequence $b_n(t) := \mathbf{Pr}(\zeta_n = t)$ is properly log-concave with respect to t . Then

$$\lim_{n \rightarrow \infty} \sigma_n \mathbf{Pr}[\zeta_n = \lfloor \sigma_n x + \mu_n \rfloor] = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

uniformly for all $x \in \mathbb{R}$.

The Lyapunov theorem is well-known; for a proof, see e.g. [17]. The Bender theorem originally appeared in [7]; see also [10], which states and proves the theorem in a form more similar to the above.

5. ACKNOWLEDGMENTS

The author thanks Alexander Barvinok for posing the problems addressed by this paper, and for many useful discussions.

REFERENCES

- [1] A. Barvinok. Brunn–Minkowski inequalities for contingency tables and integer flows. *Advances in Mathematics*, 211:105–122, 2007.
- [2] A. Barvinok. Enumerating contingency tables via random permanents. *Combinatorics, Probability, and Computing*, 17:1–19, 2008.
- [3] A. Barvinok. Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes. *International Mathematics Research Notices*, pages 348–385, 2009.
- [4] A. Barvinok. On the number of matrices and a random matrix with prescribed row and column sums and 0–1 entries. *Advances in Mathematics*, 224:316–339, 2010.
- [5] A. Barvinok. What does a random contingency table look like? *Combinatorics, Probability, and Computing*, 19:517–539, 2010.
- [6] A. Barvinok and J. Hartigan. Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes. *Advances in Applied Mathematics*, 45:252–289, 2010.
- [7] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory, Series A*, 15:91–111, 1973.
- [8] P. Diaconis and B. Efron. Testing for independence in a two-way table: new interpretations of the chi-square statistic. *Annals of Statistics*, 13:845–874, 1985.
- [9] P. Diaconis and A. Gangolli. Rectangular arrays with fixed margins. In *IMA Volumes in Mathematics and its Applications 72: Discrete Probability and Algorithms*, pages 15–41. Springer-Verlag, 1995.
- [10] K. Engel. *Sperner Theory*. Cambridge University Press, Cambridge, 1997.
- [11] L. Euler. De evolutione potestatis polynomialis cuiuscunque $(1 + x + x^2 + x^3 + x^4 + \text{etc.})^n$. *Nova Acta Academiae Scientiarum Imperialis Petropolitinae*, 12:47–57, 1801. Translation available at <http://arxiv.org/abs/math.H0/0505425>.
- [12] I.J. Good. On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Annals of Statistics*, 4:1159–1189, 1976.
- [13] S. Guiasu and A. Shenitzer. The principle of maximum entropy. *The Mathematical Intelligencer*, 7:42–48, 1985.
- [14] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.
- [15] E. T. Jaynes. Information theory and statistical mechanics II. *Physical Review*, 108:171–190, 1957.
- [16] A. Shapiro. Bounds on the number of integer points in a polytope via concentration estimates. Preprint (2010), available at <http://arxiv.org/abs/1011.6252>.
- [17] A. V. Skorokhod. *Basic Principles and Applications of Probability Theory*. Springer, Heidelberg, 2005.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109-1043, USA